HAL open science

Some results on order statistics generated by two simulation methods

N. Balakrishnan, E. Cramer, K. Davies

To cite this version:

N. Balakrishnan, E. Cramer, K. Davies. Some results on order statistics generated by two simulation methods. Statistics and Probability Letters, 2009, 79 (17), pp.1847. 10.1016/j.spl.2009.05.011 . hal-00563099

HAL Id: hal-00563099

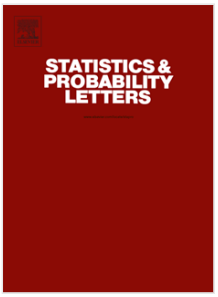
https://hal.science/hal-00563099v1

Submitted on 4 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript



Some results on order statistics generated by two simulation methods

N. Balakrishnan, E. Cramer, K. Davies

PII: S0167-7152(09)00191-6
DOI: 10.1016/j.spl.2009.05.011
Reference: STAPRO 5426

To appear in: *Statistics and Probability Letters*

Received date: 8 April 2009
Accepted date: 18 May 2009

Please cite this article as: Balakrishnan, N., Cramer, E., Davies, K., Some results on order statistics generated by two simulation methods. *Statistics and Probability Letters* (2009), doi:10.1016/j.spl.2009.05.011

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Some results on order statistics generated by two simulation methods

N. Balakrishnan[*], **E. Cramer**[+] and **K. Davies**[♯]

[*] Dept. of Mathematics and Statistics, McMaster University, Ontario, Canada L8S 4K1

[+] Institute of Statistics, RWTH Aachen University, D-52056 Aachen, Germany

[♯] Department of Statistics, University of Manitoba, Manitoba, Canada R3T 2N2

Abstract

In this paper, we consider joint distributions of order statistics generated by two simulation methods. By using these distributions, we study the nature of dependence and exceedence probabilities between them.

Keyword and Phrases: Order statistics, simulation methods, joint distribution, positive quadrant dependence, exceedence probabilities, copula, precedence-type tests.

1 Introduction

Let $U_1, \cdots, U_n$ be iid uniform random variables, and let us denote the corresponding r-th order statistic by $U_{r:n}$ and its cumulative distribution function (cdf) by $F_{r:n}$. Then, for simulating $U_{r:n}$, we have two options. First is simply to order $U_1, \cdots, U_n$ and take the r-th smallest value. Second is to take the first uniform random variable U_1 and apply a quantile transformation to define

$$U^*_{r:n} = F^{-1}_{r:n}(U_1);$$

see Nagaraja (1979) and Arnold et al. (2008). Evidently, the random variables $U_{r:n}$ and $U^*_{r:n}$ have the same distribution, but the generation process is clearly different. Moreover, they are dependent since both definitions involve a common random variable U_1.

In this paper, we are interested in the nature of dependence between $U_{r:n}$ and $U^*_{r:n}$. In Section 2, we derive the joint distribution of $U_{r:n}$ and $U^*_{r:n}$, and show that they are positive quadrant dependent. In Section 3, we derive explicit expressions for exceedence probabilities of the form $\pi^*_{r:n} = P(U^*_{r:n} \leq U_{r:n})$ and establish some associated properties. In Section 4, we present some asymptotic approximations to exceedence probabilities.

In the subsequent derivations, we make use of the following relations for cdf's of uniform order statistics, which can be found in David and Shu (1978) (see also Arnold et al. 2008, p. 40) :

$$F_{r-1:n-1}(t) = F_{r:n-1}(t) + \binom{n-1}{r-1} t^{r-1}(1-t)^{n-r} \tag{1.1}$$

ACCEPTED MANUSCRIPT

$$
\begin{aligned}
&= F_{r:n-1}(t) + \frac{1}{n} f_{r:n}(t), \quad t \in [0,1], \\
F_{r:n}(t) &= F_{r:n-1}(t) + \binom{n-1}{r-1} t^r (1-t)^{n-r} \qquad (1.2) \\
&= F_{r:n-1}(t) + \frac{1}{n} t f_{r:n}(t), \quad t \in [0,1].
\end{aligned}
$$

2 Joint distribution of $U_{r:n}$ and $U^*_{r:n}$

We shall denote the order statistics of the sample $U_2, \cdots, U_n$ by $V_{r:n-1}$ for $1 \le r \le n-1$. Note that the order statistics $V_{1:n-1}, \cdots, V_{n-1:n-1}$ are independent of U_1. For convenience, let us also define $F_{n:n-1}(x) = 0 \ \ \forall\, x \in [0,1]$.

Theorem 2.1 *The joint cdf of $U^*_{r:n}$ and $U_{r:n}$ is given by*

$$
\begin{aligned}
F^*_{r:n}(s,t) &= P(U^*_{r:n} \le s, U_{r:n} \le t) \\
&= \begin{cases} F_{r-1:n-1}(t) F_{r:n}(s), & t \ge F_{r:n}(s) \\ F_{r:n-1}(t) F_{r:n}(s) + \frac{1}{n} t f_{r:n}(t), & t < F_{r:n}(s) \end{cases}, \quad 0 \le s, t \le 1. \qquad (2.1)
\end{aligned}
$$

Proof The r-th order statistic $U_{r:n}$ in the sample $U_1, \cdots, U_n$ is related to the order statistics in the subsample $U_2, \cdots, U_n$ in the following way:

$$
U_{r:n} = \begin{cases} V_{r-1:n-1}, & U_1 < V_{r-1:n-1}, \\ U_1, & V_{r-1:n-1} \le U_1 < V_{r:n-1}, \\ V_{r:n-1}, & V_{r:n-1} \le U_1 \end{cases}.
$$

So, for $0 \le s, t \le 1$, the joint cdf can be expressed as

$$
\begin{aligned}
P(U^*_{r:n} \le s, U_{r:n} \le t) &= P(U_1 \le F_{r:n}(s), U_{r:n} \le t) \\
&= P(U_1 \le F_{r:n}(s), U_{r:n} \le t, U_1 < V_{r-1:n-1}) \\
&\quad + P(U_1 \le F_{r:n}(s), U_{r:n} \le t, V_{r-1:n-1} \le U_1 < V_{r:n-1}) \\
&\quad + P(U_1 \le F_{r:n}(s), U_{r:n} \le t, V_{r:n-1} \le U_1) \\
&= P(U_1 \le F_{r:n}(s), V_{r-1:n-1} \le t, U_1 < V_{r-1:n-1}) \\
&\quad + P(U_1 \le F_{r:n}(s), U_1 \le t, V_{r-1:n-1} \le U_1 < V_{r:n-1}) \\
&\quad + P(U_1 \le F_{r:n}(s), V_{r:n-1} \le t, V_{r:n-1} \le U_1) \\
&= I_1 + I_2 + I_3, \quad \text{say}\ .
\end{aligned}
$$

Now, we obtain the following representations for I_1, I_2 and I_3:

$$
\begin{aligned}
I_1 &= \int_0^{F_{r:n}(s)} P(V_{r-1:n-1} \le t, u < V_{r-1:n-1}) du \\
&= \int_0^{\min\{F_{r:n}(s),t\}} \Big\{ F_{r-1:n-1}(t) - F_{r-1:n-1}(u) \Big\} du \\
&= F_{r-1:n-1}(t) \min\{F_{r:n}(s), t\} - \int_0^{\min\{F_{r:n}(s),t\}} F_{r-1:n-1}(u) du, \\
I_2 &= \int_0^{\min\{F_{r:n}(s),t\}} P(V_{r-1:n-1} \le u < V_{r:n-1}) du \\
&= \int_0^{\min\{F_{r:n}(s),t\}} \binom{n-1}{r-1} u^{r-1} (1-u)^{n-r} du \\
&\overset{(1.1)}{=} \int_0^{\min\{F_{r:n}(s),t\}} \Big\{ F_{r-1:n-1}(u) - F_{r:n-1}(u) \Big\} du,
\end{aligned}
$$

ACCEPTED MANUSCRIPT

$$
\begin{aligned}
I_3 &= \int_0^{F_{r:n}(s)} P(V_{r:n-1} \le \min\{t,u\})du \\
&= \int_0^{F_{r:n}(s)} F_{r:n-1}(\min\{t,u\})du \\
&= \int_0^{\min\{F_{r:n}(s),t\}} F_{r:n-1}(u)du + \int_{\min\{F_{r:n}(s),t\}}^{F_{r:n}(s)} F_{r:n-1}(t)du \\
&= \int_0^{\min\{F_{r:n}(s),t\}} F_{r:n-1}(u)du + F_{r:n-1}(t)\Big[F_{r:n}(s) - \min\{F_{r:n}(s),t\}\Big].
\end{aligned}
$$

Summing I_2 and I_3, we obtain

$$
I_2 + I_3 = \int_0^{\min\{F_{r:n}(s),t\}} F_{r-1:n-1}(u)du + F_{r:n-1}(t)\Big[F_{r:n}(s) - \min\{F_{r:n}(s),t\}\Big]
$$

so that we finally find

$$
I_1 + I_2 + I_3 = F_{r-1:n-1}(t)\min\{F_{r:n}(s),t\} + F_{r:n-1}(t)[F_{r:n}(s) - \min\{F_{r:n}(s),t\}].
$$

Thence, we have

$$
\begin{aligned}
P(U^*_{r:n} \le s, U_{r:n} \le t) &= \begin{cases} F_{r-1:n-1}(t)F_{r:n}(s), & t \ge F_{r:n}(s) \\ F_{r-1:n-1}(t)t + F_{r:n-1}(t)[F_{r:n}(s) - t], & t < F_{r:n}(s) \end{cases} \\
&= \begin{cases} F_{r-1:n-1}(t)F_{r:n}(s), & t \ge F_{r:n}(s) \\ F_{r:n-1}(t)F_{r:n}(s) + t[F_{r-1:n-1}(t) - F_{r:n-1}(t)], & t < F_{r:n}(s) \end{cases} \\
&= \begin{cases} F_{r-1:n-1}(t)F_{r:n}(s), & t \ge F_{r:n}(s) \\ F_{r:n-1}(t)F_{r:n}(s) + \binom{n-1}{r-1}t^r(1-t)^{n-r}, & t < F_{r:n}(s) \end{cases} \\
&= \begin{cases} F_{r-1:n-1}(t)F_{r:n}(s), & t \ge F_{r:n}(s) \\ F_{r:n-1}(t)F_{r:n}(s) + \frac{1}{n}tf_{r:n}(t), & t < F_{r:n}(s) \end{cases}.
\end{aligned}
$$

Hence, the theorem. ■

Remark 2.2 *Noting that $U^*_{n:n} = U_1^{1/n}$, for $r = n$, (2.1) yields*

$$
P(U_1^{1/n} \le s, U_{n:n} \le t) = \begin{cases} t^{n-1}s^n, & t \ge s^n \\ t^n, & t < s^n \end{cases}, \quad 0 \le s,t \le 1,
$$

which is an expression given earlier by Nagaraja (1979); see also Arnold et al. (2008, p. 106).

In particular, we find from the above representation the copula of $U^*_{r:n}$ and $U_{r:n}$. Since both marginal cdf's are given by $F_{r:n}$ and so, with $s = F^{-1}_{r:n}(u)$ and $t = F^{-1}_{r:n}(v)$, we find

$$
\begin{aligned}
C(u,v) &= P(U_1 \le u, U_{r:n} \le F^{-1}_{r:n}(v)) \\
&= \begin{cases} F_{r-1:n-1}(F^{-1}_{r:n}(v))u, & v \ge F_{r:n}(u) \\ F_{r:n-1}(F^{-1}_{r:n}(v))u + \frac{1}{n}F^{-1}_{r:n}(v)f_{r:n}(F^{-1}_{r:n}(v)), & v < F_{r:n}(u) \end{cases}
\end{aligned}
$$

to be the copula of $U^*_{r:n}$ and $U_{r:n}$.

Theorem 2.3 *$U^*_{r:n}$ and $U_{r:n}$ are positive quadrant dependent, and so they are positively correlated, in particular.*

ACCEPTED MANUSCRIPT

Proof $U^*_{r:n}$ and $U_{r:n}$ are positive quadrant dependent (PQD) if

$$P(U^*_{r:n} \le s, U_{r:n} \le t) \ge P(U^*_{r:n} \le s)Pr(U_{r:n} \le t) = F_{r:n}(s)F_{r:n}(t),\ 0 \le s, t \le 1.$$

By using the result (Shaked and Shanthikumar 2007, p. 277)

$$U_{i:n} \le_{st} U_{j:m} \quad \Longleftrightarrow \quad i \le j \text{ and } n - i \ge m - j,$$

we find $U_{r-1:n-1} \le_{st} U_{r:n}$ so that $F_{r-1:n-1} \ge F_{r:n}$. Therefore, we get from (2.1), for $t \ge F_{r:n}(s)$,

$$F^*_{r:n}(s,t) = F_{r-1:n-1}(t)F_{r:n}(s) \ge F_{r:n}(s)F_{r:n}(t).$$

Let $t < F_{r:n}(s)$; then

$$F^*_{r:n}(s,t) = F_{r:n-1}(t)F_{r:n}(s) + \frac{1}{n}tf_{r:n}(t) \ge F_{r:n}(s)F_{r:n}(t)$$

iff

$$F_{r:n}(s)\Big\{F_{r:n-1}(t) - F_{r:n}(t)\Big\} + \frac{1}{n}tf_{r:n}(t) \ge 0. \tag{2.2}$$

Now, from (1.2), we have $F_{r:n-1}(t) - F_{r:n}(t) = -\frac{1}{n}tf_{r:n}(t)$, and so (2.2) is equivalent to

$$\frac{1}{n}tf_{r:n}(t)[1 - F_{r:n}(s)] \ge 0.$$

This obviously holds, which implies the positive quadrant dependence of $U^*_{r:n}$ and $U_{r:n}$. The positive correlation then readily follows. ■

Using Hoeffding's expression for the covariance and $\int_0^1 F_{r:n}(s)ds = \frac{n-r+1}{n+1}$, we find

$$\begin{aligned}
\frac{1}{12}\mathrm{Cov}(U^*_{r:n}, U_{r:n}) &= \int_0^1 \int_0^{F_{r:n}(s)} \left\{F_{r:n-1}(t)F_{r:n}(s) + \frac{1}{n}tf_{r:n}(t)\right\} dtds \\
&\quad + \int_0^1 \int_{F_{r:n}(s)}^1 F_{r-1:n-1}(t)F_{r:n}(s)dtds - \left(\int_0^1 F_{r:n}(s)ds\right)^2 \\
&= \int_0^1 \int_0^{F_{r:n}(s)} \left[\{F_{r:n-1}(t) - F_{r-1:n-1}(t)\}\, F_{r:n}(s) + \frac{1}{n}tf_{r:n}(t)\right] dtds \\
&\quad + \int_0^1 \int_0^1 F_{r-1:n-1}(t)F_{r:n}(s)dtds - \left(\frac{n-r+1}{n+1}\right)^2 \\
&= \frac{1}{n}\int_0^1 \int_0^{F_{r:n}(s)} f_{r:n}(t)\,\{t - F_{r:n}(s)\}\, dtds \\
&\quad + \frac{(n-r+1)^2}{n(n+1)} - \left(\frac{n-r+1}{n+1}\right)^2 \\
&= \frac{1}{n}\int_0^1 \int_0^{F_{r:n}(s)} f_{r:n}(t)\,\{t - F_{r:n}(s)\}\, dtds + \frac{(n-r+1)^2}{n(n+1)^2} \\
&= -\frac{1}{n}\int_0^1 \int_0^{F_{r:n}(s)} F_{r:n}(t)dtds + \frac{(n-r+1)^2}{n(n+1)^2}.
\end{aligned}$$

Since $\mathrm{Cov}(U^*_{r:n}, U_{r:n}) \ge 0$ by the PQD-property established in Theorem 2.3, we obtain the following bounds for the covariance:

$$0 \le \mathrm{Cov}(U^*_{r:n}, U_{r:n}) \le 12\frac{(n-r+1)^2}{n(n+1)^2} \le \frac{12}{n},$$

which readily reveals that the covariance tends to zero as $n \to \infty$.

ACCEPTED MANUSCRIPT

3 Exceedence probabilities

In this section, we obtain an expression for the exceedence probabilities $\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n})$ for $r = 1, \cdots, n$. First, we get the following decomposition of $\pi^*_{r:n}$:

$$\begin{aligned}\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n}) &= P(F^{-1}_{r:n}(U_1) \le U_{r:n}) \\ &= P(F^{-1}_{r:n}(U_1) \le V_{r-1:n-1}, U_1 < V_{r-1:n-1}) \\ &\quad + P(F^{-1}_{r:n}(U_1) \le U_1, V_{r-1:n-1} \le U_1 < V_{r:n-1}) \\ &\quad + P(F^{-1}_{r:n}(U_1) \le V_{r:n-1}, V_{r:n-1} \le U_1) \\ &= J_1 + J_2 + J_3, \quad \text{say .} \end{aligned} \tag{3.1}$$

It turns out that the solution of the equation

$$t = F_{r:n}(t), \quad t \in (0,1), \tag{3.2}$$

is important in the calculation of the probabilities J_1, J_2 and J_3. $F_{r:n}$ is an S-shaped function which is unimodal, as seen in Figure 3.1. It is first convex and then concave with $F_{r:n}(0) = 0$ and $F_{r:n}(1) = 1$. For $1 < r < n$, $F_{r:n}$ has exactly one fixpoint in the interval $(0,1)$, i.e., there exists exactly one solution $y_{r:n}$ of (3.2). Observe that (3.2) has no solution for $r \in \{1, n\}$. In these cases, let us define for convenience $y_{1:n} = 0$ and $y_{n:n} = 1$. Geometrically, $y_{r:n}$ is the intersection point where the graph of $F_{r:n}$ intersects the line from $(0,0)$ to $(1,1)$. Moreover,

$$F_{r:n}(t) < t \quad \text{for} \quad t \in (0, y_{r:n}) \qquad \text{and} \qquad F_{r:n}(t) > t \quad \text{for} \quad t \in (y_{r:n}, 1). \tag{3.3}$$

For a given n, the solutions $y_{r:n} \in [0,1]$ of Eq. (3.2), for $r = 1, \cdots, n$, have the following properties:

Lemma 3.1 *Let n be fixed. Then:*

(i) $y_{r:n}$ is strictly increasing in r;

(ii) $y_{n-r+1:n} = 1 - y_{r:n}$, $r = 1, \cdots, n$;

(iii) For n odd and $r = \frac{n+1}{2}$, we have $y_{r:n} = \frac{1}{2}$;

(iv) For $1 \le r < \frac{n+1}{2}$, we have $y_{r:n} < \frac{1}{2}$;

(v) For $\frac{n+1}{2} < r \le n$, we have $y_{r:n} > \frac{1}{2}$.

Proof (i) Suppose that $y_{r:n} < y_{r-1:n}$. Then, we conclude from (3.3) that

$$y_{r-1:n} < F_{r:n}(y_{r-1:n}).$$

The cdf's $F_{r:n}$ are strictly ordered with respect to r, i.e., for fixed n, $F_{r:n}(t) < F_{r-1:n}(t)$, $t \in [0,1]$ (see (1.1)). Therefore, $F_{r:n}(y_{r-1:n}) < F_{r-1:n}(y_{r-1:n}) = y_{r-1:n}$ which results in a contradiction. Thus, $y_{r-1:n} \le y_{r:n}$, $1 \le r \le n$. Moreover, equality is not possible. For $r = 1$, we have $y_{1:n} = 0 < y_{2:n}$. Assuming that $r \ge 2$, and that $y_{r:n} = y_{r-1:n}$, we deduce from (1.1) that

$$0 = F_{r:n}(y_{r:n}) - F_{r-1:n}(y_{r:n}) = -\frac{1}{n} f_{r:n+1}(y_{r:n}) < 0.$$

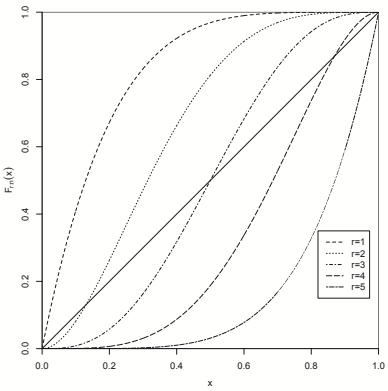
Thus, we obtain a contradiction once again, and so $y_{r:n}$ is strictly increasing in r.

(ii) Using the fact that $F_{r:n}(t) = 1 - F_{n-r+1:n}(1-t)$, $t \in [0,1]$, we have

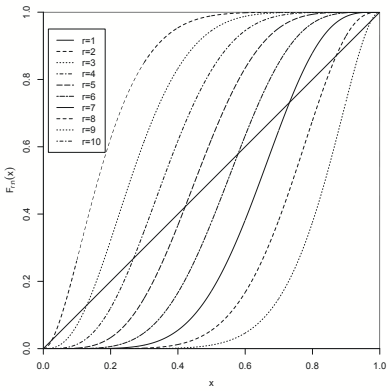
$$F_{r:n}(t) = t \iff 1 - t = F_{n-r+1:n}(1-t).$$

Therefore, α is a fixpoint of $F_{r:n}$ iff $1 - \alpha$ is a fixpoint of $F_{n-r+1:n}$.

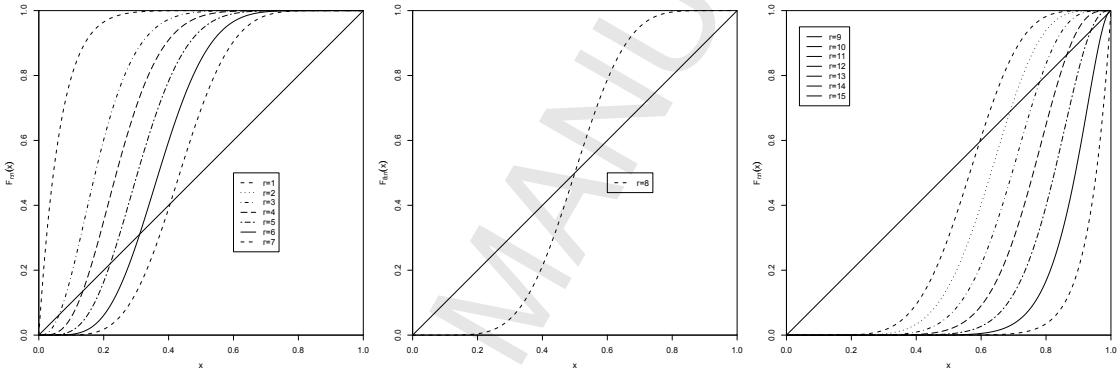
ACCEPTED MANUSCRIPT



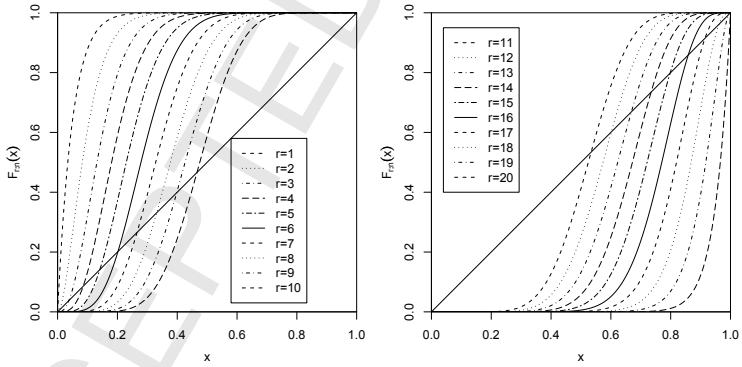
(a) $n = 5$



(b) $n = 10$



(c) $n = 15$



(d) $n = 20$

Figure 3.1: Plots of cumulative distribution functions of order statistics for $n = 5, 10, 15, 20$.

ACCEPTED MANUSCRIPT

n	r	2	3	4	5	6	7	8	9	10
5		0.1311	0.5000							
10		0.0255	0.1294	0.2683	0.4214					
15		0.0104	0.0621	0.1357	0.2203	0.3108	0.4047	0.5000		
20		0.0056	0.0376	0.0854	0.1412	0.2018	0.2655	0.3313	0.3983	0.4660

Table 3.1: Intersection points $y_{r:n}$ for $n = 5, 10, 15, 20$.

(iii) From (ii), we conclude for n odd and $r = \frac{n+1}{2}$, that $n - r + 1 = r$. Therefore, $y_{r:n}$ and $1 - y_{r:n}$ are fixpoints. Since we have exactly one fixpoint, we conclude that $y_{r:n} = \frac{1}{2}$.

(iv) Assertions (iv) and (v) follow immediately from (i) and (ii). ■

These properties of $y_{r:n}$, $1 \le r \le n$, are illustrated in Figure 3.1. Table 3.1 gives intersection points $y_{r:n}$ for $n = 5, 10, 15, 20$. The values are given only for $r = 2, \ldots, \lfloor (n+1)/2 \rfloor$ since $y_{n-r+1:n} = 1 - y_{r:n}$ (see Part (ii) of Lemma 3.1). The computation of $y_{r:n}$ is difficult for large values of n, but we conjecture the following bounds. Notice that $\frac{r-1}{n-1}$ is the change point of $F_{r:n}$.

Conjecture 3.2 *Let $2 \le r \le \lfloor (n+1)/2 \rfloor$ and $w_{r:n}$ be the smallest change point of $f_{r:n}$ given by*

$$w_{r:n} = \frac{r-1}{n-1} - \frac{\sqrt{2r^2 - r^2 n - rn - 2r + rn^2 - n^2 + 2n}}{(n-1)(n-2)}.$$

Then,

$$w_{r:n} \le y_{r:n} \le \frac{r-1}{n-1}.$$

Moreover, for $(r_n)_n$ with $\lim_{n\to\infty} \frac{r_n}{n} = \alpha \in (0,1)$, we have

$$\lim_{n\to\infty} \sqrt{n}\left(\frac{r_n}{n} - y_{r_n:n}\right) = \sqrt{\alpha(1-\alpha)}.$$

Now, we are ready to establish the following theorem which gives a simple expression for the exceedence probability $\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n})$.

Theorem 3.3 *The exceedence probability $\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n})$, $1 \le r \le n$, is given by*

$$\pi^*_{r:n} = \frac{1}{2} + \frac{1}{2}\frac{\left\{\binom{n-1}{r-1}\right\}^2}{\binom{2n-1}{2r-1}} - \frac{\binom{2r-2}{r-1}\binom{2n-2r}{n-r}}{\binom{2n-1}{n-1}} F_{2r-1:2n-1}(y_{r:n}) \tag{3.4}$$

$$= \frac{1}{2} + \frac{\binom{2r-2}{r-1}\binom{2n-2r}{n-r}}{\binom{2n-1}{n-1}}\left\{\frac{2r-1}{2n} - F_{2r-1:2n-1}(y_{r:n})\right\}. \tag{3.5}$$

Proof Suppose $1 < r < n$. Then, we have from (3.1),

$$\begin{aligned} J_1 &= \int_0^1 P(U_1 \le F_{r:n}(u), U_1 \le u) f_{r-1:n-1}(u)du \\ &= \int_0^1 \min\{F_{r:n}(u), u\} f_{r-1:n-1}(u)du \\ &= \int_0^{y_{r:n}} F_{r:n}(u) f_{r-1:n-1}(u)du + \int_{y_{r:n}}^1 u f_{r-1:n-1}(u)du \end{aligned}$$

ACCEPTED MANUSCRIPT

$$
\begin{aligned}
J_2 &= \int_0^1 P(u \le F_{r:n}(u), V_{r-1:n-1} \le u < V_{r:n-1})du \\
&= \int_{y_{r:n}}^1 P(V_{r-1:n-1} \le u < V_{r:n-1})du \\
&= \frac{1}{n}\int_{y_{r:n}}^1 f_{r:n}(u)du = \frac{1}{n}\left\{1 - F_{r:n}(y_{r:n})\right\} = \frac{1}{n}(1 - y_{r:n}), \\
J_3 &= \int_0^1 P(U_1 \le F_{r:n}(u), u \le U_1) f_{r:n-1}(u)du \\
&= \int_{y_{r:n}}^1 P(u \le U_1 \le F_{r:n}(u)) f_{r:n-1}(u)du \\
&= \int_{y_{r:n}}^1 \left\{F_{r:n}(u) - u\right\} f_{r:n-1}(u)du.
\end{aligned}
$$

From (1.1), we have

$$
f_{r-1:n-1}(t) = f_{r:n-1}(t) + \frac{1}{n} f'_{r:n}(t), \quad t \in (0,1). \tag{3.6}
$$

So, by partial integration, we obtain

$$
\begin{aligned}
\int_{y_{r:n}}^1 u f_{r-1:n-1}(u)du - \int_{y_{r:n}}^1 u f_{r:n-1}(u)du &= \frac{1}{n}\int_{y_{r:n}}^1 u f'_{r:n}(u)du \\
&= -\frac{1}{n} y_{r:n} f_{r:n}(y_{r:n}) - \frac{1}{n}(1 - y_{r:n}).
\end{aligned}
$$

Moreover, by (3.6), we have

$$
\begin{aligned}
&\int_0^{y_{r:n}} F_{r:n}(u) f_{r-1:n-1}(u)du - \int_0^{y_{r:n}} F_{r:n}(u) f_{r:n-1}(u)du \\
&\quad = \frac{1}{n}\int_0^{y_{r:n}} F_{r:n}(u) f'_{r:n}(u)du \\
&\quad = \frac{1}{n}\Big[F_{r:n}(u) f_{r:n}(u)\Big]_0^{y_{r:n}} - \frac{1}{n}\int_0^{y_{r:n}} f^2_{r:n}(u)du \\
&\quad = \frac{1}{n} y_{r:n} f_{r:n}(y_{r:n}) - \frac{1}{n}\int_0^{y_{r:n}} f^2_{r:n}(u)du.
\end{aligned}
$$

Thus, summing up J_1, J_2 and J_3, we arrive at

$$
\begin{aligned}
P(U^*_{r:n} \le U_{r:n}) &= \int_0^1 F_{r:n}(u) f_{r:n-1}(u)du - \frac{1}{n}\int_0^{y_{r:n}} f^2_{r:n}(u)du \\
&= P(U_{r:n} \le V_{r:n-1}) - \frac{1}{n}\int_0^{y_{r:n}} f^2_{r:n}(u)du.
\end{aligned}
$$

Since,

$$
\frac{1}{n}\int_0^{y_{r:n}} f^2_{r:n}(u)du = \frac{\binom{2r-2}{r-1}\binom{2n-2r}{n-r}}{\binom{2n-1}{n-1}} F_{2r-1:2n-1}(y_{r:n}),
$$

we get

$$
P(U^*_{r:n} \le U_{r:n}) = P(U_{r:n} \le V_{r:n-1}) - \frac{\binom{2r-2}{r-1}\binom{2n-2r}{n-r}}{\binom{2n-1}{n-1}} F_{2r-1:2n-1}(y_{r:n}). \tag{3.7}
$$

ACCEPTED MANUSCRIPT

The expression in (3.7) can be evaluated by using (1.2). As

$$\begin{aligned}
P(U_{r:n} \le V_{r:n-1}) &= \int_0^1 F_{r:n}(t) f_{r:n-1}(t)dt \\
&= \int_0^1 F_{r:n-1}(t) f_{r:n-1}(t)dt + \int_0^1 \binom{n-1}{r-1} t^r (1-t)^{n-r} f_{r:n-1}(t)dt \\
&= P(U_{r:n-1} \le V_{r:n-1}) + \left\{\binom{n-1}{r-1}\right\}^2 (n-r) \int_0^1 t^{2r-1}(1-t)^{2n-2r-1}dt \\
&= \frac{1}{2} + \frac{1}{2}\frac{\binom{n-1}{r-1}^2}{\binom{2n-1}{2r-1}},
\end{aligned}$$

the theorem follows. ■

Remark 3.4 *For $r = n$, we deduce*

$$\begin{aligned}
P(U_1^{1/n} \le U_{n:n}) &= \int_0^1 P(U_1 \le \max\{U_1^n, u^n\})(n-1)u^{n-2}du = \int_0^1 u^n(n-1)u^{n-2}du \\
&= (n-1)\int_0^1 u^{2n-2}du = \frac{n-1}{2n-1},
\end{aligned}$$

which has been derived by Nagaraja (1979). For $r = 1$, with $V_j = 1 - U_j$, $1 \le j \le n$, we find

$$P(1-(1-U_1)^{1/n} \le U_{1:n}) = P(1 - U_{1:n} \le (1-U_1)^{1/n}) = P(V_{n:n} \le V_1^{1/n}) = \frac{n}{2n-1}.$$

Remark 3.5 *By applying a result known for precedence-type tests (see Balakrishnan and Ng 2006, p. 34/35), we find*

$$P(U_{r:n} \le V_{r:n-1}) = \sum_{j=r}^{n} \frac{\binom{j+r-1}{r-1}\binom{2n-j-r-1}{n-j}}{\binom{2n-1}{n-1}}.$$

A scatterplot of 2000 simulated random pairs $(U_{6:15}, U^*_{6:15})$ is depicted in Figure 3.2. For establishing some properties of the exceedence probabilities $\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n})$, we need the following lemma which gives some insight into the support of the joint distribution $F^*_{r:n}$. In particular, it shows that the set

$$G = \{(F^{-1}(u), u) \,|\, u \in [0,1]\}$$

has non-zero measure. Consequently, $F^*_{r:n}$ is not absolutely continuous w.r.t. the two-dimensional Lebesgue measure showing that it has no Lebesgue density function.

Lemma 3.6 *For any $1 \le r \le n$,*

$$P(U^*_{r:n} = U_{r:n}) = 0.$$

Moreover,

$$P\Big((U^*_{r:n}, U_{r:n}) \in G\Big) = \frac{1}{n},$$

and, for $G_> = \{(t,u) \,|\, t > F^{-1}_{r:n}(u), u \in [0,1]\}$ and $G_< = \{(t,u) \,|\, t < F^{-1}_{r:n}(u), u \in [0,1]\}$

$$P\Big((U^*_{r:n}, U_{r:n}) \in G_>\Big) = 1 - \frac{r}{n} \quad \text{and} \quad P\Big((U^*_{r:n}, U_{r:n}) \in G_<\Big) = \frac{r-1}{n}.$$

ACCEPTED MANUSCRIPT

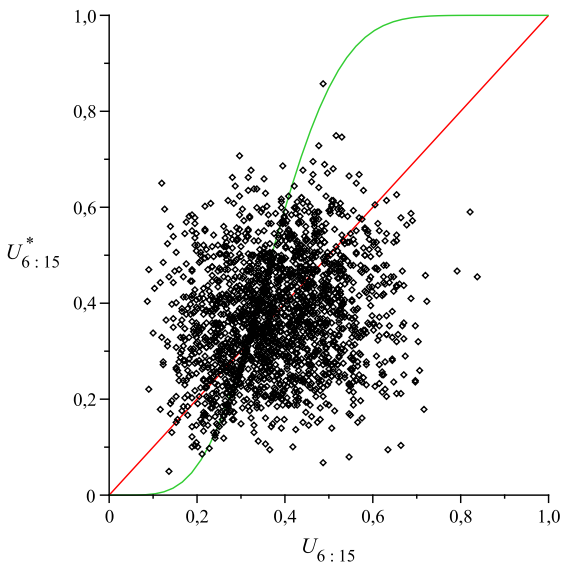


Figure 3.2: Scatterplot of 2000 simulated random pairs $(U_{6:15}, U^*_{6:15})$.

Proof As above, we can write

$$\begin{aligned} P(U^*_{r:n} = U_{r:n}) &= P\Big(F^{-1}_{r:n}(U_1) = V_{r-1:n-1}, U_1 < V_{r-1:n-1}\Big) \\ &\quad + P\Big(F^{-1}_{r:n}(U_1) = U_1, V_{r-1:n-1} \le U_1 < V_{r:n-1}\Big) \\ &\quad + P\Big(F^{-1}_{r:n}(U_1) = V_{r:n-1}, V_{r:n-1} \le U_1\Big) \\ &\le P\Big(F^{-1}_{r:n}(U_1) = V_{r-1:n-1}\Big) + P\Big(F^{-1}_{r:n}(U_1) = U_1\Big) \\ &\quad + P\Big(F^{-1}_{r:n}(U_1) = V_{r:n-1}\Big) = 0. \end{aligned}$$

The second result can be seen as follows:

$$P\Big((U^*_{r:n}, U_{r:n}) \in G\Big) = P\Big((F^{-1}_{r:n}(U_1), U_{r:n}) \in G\Big) = P\Big(U_1 = U_{r:n}\Big) = \frac{1}{n}.$$

Finally, we find

$$\begin{aligned} P\Big((U^*_{r:n}, U_{r:n}) \in G_>\Big) &= P\Big(U^*_{r:n} > F^{-1}_{r:n}(U_{r:n})\Big) = P\Big(U_1 > U_{r:n}\Big) \\ &= P\Big(U_1 > V_{r:n-1}\Big) = \int_0^1 F_{r:n-1}(t)dt = \frac{n-r}{n} = 1 - \frac{r}{n}. \end{aligned}$$

In an analogous manner, we find $P\Big((U^*_{r:n}, U_{r:n}) \in G_>\Big) = \frac{r-1}{n}$. ■

By combining the above results, we obtain the following theorem.

Theorem 3.7 *For any $1 \le r \le n$, we have*

$$\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n}) = 1 - P(U^*_{n-r+1:n} \le U_{n-r+1:n}) = 1 - \pi^*_{n-r+1:n}.$$

In particular, for n odd and $r = \frac{n+1}{2}$, we have

$$P(U^*_{r:n} \le U_{r:n}) = \frac{1}{2},$$

ACCEPTED MANUSCRIPT

n	r	1	2	3	4	5	6	7	8	9	10
5		0.5556	0.5625	0.5000							
10		0.5263	0.5385	0.5359	0.5244	0.5086					
15		0.5172	0.5259	0.5283	0.5263	0.5216	0.5152	0.5078	0.5000		
20		0.5128	0.5194	0.5224	0.5227	0.5213	0.5187	0.5153	0.5113	0.5069	0.5023

Table 3.2: Exceedence probabilities $\pi^*_{r:n}$ for $n = 5, 10, 15, 20$.

so that in the case of the median the probability equals $\frac{1}{2}$.

Proof According to Lemma 3.6, $P\Big(F^{-1}_{r:n}(U_1) = U_{r:n}\Big) = 0$ so that we do not have to take care of equality. Then, due to the relations $U_{r:n} = 1 - (1-U)_{n-r+1:n}$ and $F_{n-r+1:n}(t) = 1 - F_{r:n}(1-t)$ for $t \in [0,1]$, we find

$$\begin{aligned}
P(U^*_{r:n} \le U_{r:n}) &= P\Big(F^{-1}_{r:n}(U_1) \le U_{r:n}\Big) \\
&= P\Big(1 - F^{-1}_{n-r+1:n}(1-U_1) \le 1 - (1-U)_{n-r+1:n}\Big) \\
&= P\Big((1-U)_{n-r+1:n} \le F^{-1}_{n-r+1:n}(1-U_1)\Big) \\
&= P\Big(U_{n-r+1:n} \le F^{-1}_{n-r+1:n}(U_1)\Big) \\
&= 1 - P\Big(U^*_{n-r+1:n} \le U_{n-r+1:n}\Big).
\end{aligned}$$

Note that both $1-U$ and U have the uniform distribution on $[0,1]$. The second assertion follows since, in this case, $n - r + 1 = \frac{n+1}{2} = r$ and so both events are the same. ■

Table 3.2 gives exceedence probabilities $\pi^*_{r:n}$ for $n = 5, 10, 15, 20$. Due to Theorem 3.7, we only present the values of $\pi^*_{r:n}$ for $r = 1, \cdots, \lfloor (n+1)/2 \rfloor$.

Conjecture 3.8 *For any* $1 \le r < \frac{n+1}{2}$, *we have*

$$\pi^*_{r:n} = P(U^*_{r:n} \le U_{r:n}) > \frac{1}{2}.$$

For $r = 1$, the result follows directly since $y_{1:n} = 0$. Let $r \ge 2$. Then, from (3.5), it is sufficient to show that

$$\frac{2r-1}{2n} \ge F_{2r-1:2n-1}(y_{r:n}), \quad 2 \le r \le \frac{n+1}{2}.$$

The latter condition is equivalent to

$$E\left(U_{2r-1:2n-1}\right) \ge F_{2r-1:2n-1}(y_{r:n}) = P(U_{2r-1:2n-1} \le y_{r:n}).$$

Finally, we conjecture that the sequence $\pi^*_{r:n}$ is increasing-decreasing.

Conjecture 3.9

$$\pi^*_{r:n} \textit{ is increasing-decreasing in } r \textit{ for } 1 \le r \le \frac{n+1}{2}.$$

ACCEPTED MANUSCRIPT

4 Asymptotics

Lemma 4.1 *Suppose $y_{r:n}$ is the unique solution of the equation $F_{r:n}(t) = t$ in the interval $(0,1)$. Moreover, let $(r_n)_n$ be a sequence such that $\lim_{n\to\infty} \frac{r_n}{n} = \alpha \in (0,1)$. Then, we have*

$$\lim_{n\to\infty} y_{r_n:n} = \alpha.$$

Proof Let $t \in (0,1)$. Suppose S_n is a binomial random variable with parameters n and t. Then, with $S_n^* = \frac{S_n - nt}{\sqrt{nt(1-t)}}$, we have

$$F_{r_n:n}(t) = P(S_n \geq r_n) = P\left(S_n^* \geq \frac{r_n - nt}{\sqrt{nt(1-t)}}\right) = P\left(\frac{1}{\sqrt{n}} S_n^* \geq \frac{\frac{r_n}{n} - t}{\sqrt{t(1-t)}}\right).$$

By central limit theorem, S_n^* is asymptotically standard normal so that $\frac{1}{\sqrt{n}} S_n^*$ converges to zero in probability. Moreover, $\lim_{n\to\infty} \frac{r_n/n - t}{\sqrt{t(1-t)}} = \frac{\alpha - t}{\sqrt{t(1-t)}}$. Hence, we obtain from Slutsky's Lemma that

$$\lim_{n\to\infty} F_{r_n:n}(t) = \begin{cases} 0, & t \leq \alpha \\ 1, & t > \alpha \end{cases}. \tag{4.1}$$

Now, for any r and n, $F_{r:n}$ is a continuous function such that the equation $F_{r:n}(t) = t$ has exactly one solution $y_{r:n}$ in the interval $(0,1)$. Let $\delta > 0$ and $\varepsilon > 0$. From (4.1), we find that there exists n_1 such that $F_{r_n:n}(z) < \delta$ for all $n \geq n_1$ and $z \in (0, \alpha - \varepsilon)$. Analogously, we find that there exists n_2 such that $F_{r_n:n}(z) > 1 - \delta$ for all $n \geq n_2$ and $z \in (\alpha + \varepsilon, 1)$. Hence, the solution of $F_{r_n:n}(t) = t$ must be contained in the interval $(\alpha - \varepsilon, \alpha + \varepsilon)$ for all $n \geq \max\{n_1, n_2\}$. Since ε and δ are chosen arbitrarily, we find that $\lim_{n\to\infty} y_{r_n:n} = \alpha$. ∎

Conjecture 4.2 *Let $(r_n)_n$ be a sequence such that $\lim_{n\to\infty} \frac{r_n}{n} = \alpha \in (0,1)$. Then,*

$$\lim_{n\to\infty} F_{2r_n-1:2n-1}(y_{r_n:n}) = v(\alpha)$$

for an appropriate function $v : [0,1] \to [0,1]$.

Theorem 4.3 *Let $(r_n)_n$ be a sequence such that $\lim_{n\to\infty} \frac{r_n}{n} = \alpha \in (0,1)$. Then,*

$$\lim_{n\to\infty} \pi^*_{r_n:n} = \frac{1}{2}.$$

Proof From (3.5), we have

$$\pi^*_{r_n:n} = \frac{1}{2} + \frac{\binom{2r_n-2}{r_n-1}\binom{2n-2r_n}{n-r_n}}{\binom{2n-1}{n-1}} \left\{\frac{2r_n - 1}{2n} - F_{2r_n-1:2n-1}(y_{r_n:n})\right\}.$$

From Stirling's formula, we have for large k

$$\binom{2k}{k} \approx \frac{2^{2k}}{\sqrt{\pi k}}$$

using which we find

$$\frac{\binom{2r_n-2}{r-1}\binom{2n-2r_n}{n-r_n}}{\binom{2n-1}{n-1}} \approx \frac{\frac{2^{2r_n-2}}{\sqrt{\pi r_n}} \cdot \frac{2^{2(n-r_n)}}{\sqrt{\pi(n-r_n)}}}{\frac{2^{2n-2}}{\sqrt{\pi n}}} = \frac{1}{\sqrt{n\pi \cdot r_n/n(1 - r_n/n)}} \approx \frac{1}{\sqrt{n\pi\alpha(1-\alpha)}}.$$

Since $F_{2r_n-1:2n-1}(y_{r_n:n}) \in [0,1]$, we obtain the desired result. ∎

ACCEPTED MANUSCRIPT

References

Arnold, B. C., Balakrishnan, N. and Nagaraja, H. N. (2008). *A First Course in Order Statistics*, Classic Edition, SIAM, Philadelphia.

Balakrishnan, N. and Ng, H. K. T. (2006). *Precedence-Type Tests and Applications,* John Wiley & Sons, Hoboken, New Jersey.

David, H. A. and Shu, V. S. (1978). Robustness of location estimators in the presence of an outlier, In *Contributions to Survey Sampling and Applied Statistics* (Ed., H. A. David), pp. 235–250, Academic Press, New York.

Nagaraja, H. N. (1979). Some relations between order statistics generated by different methods, *Commun. Statist. – Simul. Comput.* **8**, 369–377.

Shaked, M. and Shanthikumar, J. G. (2007). *Stochastic Orders*, Springer-Verlag, New York.

ACCEPTED MANUSCRIPT